AN ANALYSIS OF SCALE-SPACE SAMPLING IN SIFT

Ives Rey-Otero[†]*, Jean-Michel Morel*[†]*, Mauricio Delbracio*[†,§]

[†]CMLA, ENS-Cachan, France
[§]ECE, Duke University, USA

ABSTRACT

The most popular image matching algorithm SIFT, introduced by D. Lowe a decade ago, has proven to be sufficiently scale invariant to be used in numerous applications. In practice, however, scale invariance may be weakened by various sources of error. The density of the sampling of the Gaussian scale-space and the level of blur in the input image are two of these sources. This article presents an empirical analysis of their impact on the extracted keypoints stability. We prove that SIFT is really scale and translation invariant only if the scale-space is significantly oversampled. We also demonstrate that the threshold on the difference of Gaussians value is inefficient for eliminating aliasing perturbations.

***Index Terms*—** SIFT, invariance, scale-space, sampling, aliasing

1. INTRODUCTION

SIFT [1, 2] is a popular image matching method extensively used in image processing and computer vision applications. SIFT relies on the extraction of keypoints and the computation of local invariant feature descriptors. The property of scale invariance is crucial. The matching of SIFT features is used in various applications such as, image stitching [3], 3D reconstruction [4] and camera calibration [5].

SIFT was proved to be theoretically scale invariant [6]. Indeed, SIFT keypoints are covariant, being the extrema of the image Gaussian scale-space [7, 8]. In practice, however, the computation of the SIFT keypoints is affected in many ways, which in turn limits the scale invariance. For instance, the extraction of continuous extrema from a discrete scale-space is a challenging task. We shall show that the solution adopted by SIFT is rudimentary and may be affected by the sampling and the noise in the input image. We prove that the blur level in the input image also limits SIFT performance. Artifacts caused by undersampling degrade the SIFT keypoint stability.

The literature on SIFT focuses on variants, alternatives and accelerations [3, 9–33]. Yet, the huge amount of citations of the SIFT articles indicates that it has become a standard and a reference in many applications. In contrast there are almost no articles discussing the SIFT settings and trying to compare SIFT with itself. By this comparison we mean the question of comparing the SIFT invariance claim with its empirical invariance, and the influence of the SIFT parameters on its own performance. On this strict subject D. Lowe's paper [2] remains the principal reference, and it seems that very few of its claims on the parameter choices of the method have undergone a serious scrutiny. This paper intends to fill in the gap for the main claim of the SIFT method, namely its scale invariance, and incidentally on its translation invariance.

In this paper we investigate the role of the SIFT parameters by means of a strict image simulation framework. This permits to control the main image and scale-space sampling parameters: initial blur, scale and space sampling, noise level. We prove that scale-space sampling has an important influence on the scale invariance and that the robust extraction of all scale-space extrema requires to significantly oversample the Gaussian scale-space. We experimentally demonstrate that the invariance is limited by the aliasing in the input image whereas large scale detections are less affected. Also, we show that the contrast threshold proposed in SIFT is ineffective to remove the unstable detections due to aliasing in the input image.

The remainder of the paper is organized as follows. Section 2 briefly presents the SIFT algorithm and details how the Gaussian scale-space is implemented. Section 3 exposes the theoretical scale invariance. With that aim in view, we explicit the camera model consistent with SIFT. The experiments in Section 4 explore the limits of SIFT numerical consistency. In particular we exhibit how the invariance property is significantly affected by the sampling of the scale-space and by the blur level in the input image. Section 5 is a conclusive discussion.

2. THE SIFT METHOD

2.1. SIFT overview

SIFT derives from scale invariance properties of the Gaussian scale-space [7,8]. The Gaussian scale-space of an initial image u is the 3D function

$$v : (\sigma, \mathbf{x}) \mapsto G_\sigma u(\mathbf{x}),$$

where $G_\sigma u(\mathbf{x})$ denotes the convolution of $u(\mathbf{x})$ with a Gaussian kernel of standard deviation σ (the scale). In this framework, the Gaussian kernel acts as an approximation of the optical blur introduced in the camera (represented by its point spread function). Among other important properties [8], the Gaussian approximation is convenient because it satisfies the semi-group property

$$G_\sigma(G_\gamma u)(\mathbf{x}) = G_{\sqrt{\sigma^2+\gamma^2}} u(\mathbf{x}). \tag{1}$$

In particular, this permits to simulate distant snapshots from closer ones. Thus, the scale-space can be seen as a stack of images, each one corresponding to a different zoom factor. Matching two images with SIFT consists in matching keypoints extracted from these two stacks.

SIFT keypoints are defined as the 3D extrema of the difference of Gaussians (DoG) scale-space. Let v be the Gaussian scale-space, the DoG is the 3D function

$$w : (\sigma, \mathbf{x}) \mapsto v(\kappa\sigma, \mathbf{x}) - v(\sigma, \mathbf{x}),$$

where $\kappa > 1$ is a parameter controlling the scale sampling density. The DoG operator can be seen as an approximation of the normalized Laplacian of the scale-space $\sigma^2 \Delta v(\sigma, \mathbf{x})$ [2, 8].

Extracting the 3D *continuous* extrema from the observed *discrete* Gaussian scale-space is a difficult task. SIFT proceeds as follows. The DoG scale-space is first scanned for discrete extrema, each voxel being compared to its 26 neighbors. Then a local quadratic model is computed around each extremum to refine the extrema position. As we will show, this rudimentary approach is significantly sensitive to scale-space sampling. To compensate this shortcoming, SIFT incorporates two filters that seek to discard the unreliable detections. Uncontrasted detections are filtered out by discarding those keypoints with a small DoG value. Keypoints lying on edges are also discarded since their location is not precise due to their intrinsic translation invariant nature. A reference keypoint orientation is computed based on the dominant gradient orientation in the keypoint surrounding. This orientation along with the keypoint coordinates are used to extract a covariant patch. Finally, the gradient orientation distribution in this patch is coded into a 128 elements feature, the so-called SIFT descriptor. We shall not discuss further the constitution of the descriptor and refer to the abundant literature [17, 18, 30, 33–35].

2.2. The architecture of the Gaussian scale-space

The Gaussian digital scale-space consists of a set of digital images with different blur levels and different sampling rates, all of them derived from the input image with assumed blur level c.

The construction of the digital scale-space begins with the computation of a *seed* image. The input image is oversampled by a factor $1/\delta_{\min}$ and filtered by a Gaussian kernel $G_{(\sigma_{\min}{}^2-c^2)^{1/2}}$ to reach the minimal level of blur $\sigma_{\min}$ and inter-pixel distance $\delta_{\min}$. The scale-space set is split into subsets where images share a common inter-pixel distance. Since in the original SIFT algorithm the sampling rate is iteratively decreased by a factor of two, these subsets are called *octaves*. Denoting by n_{spo} the number of scales per octave, each image at each octave has a different blur level. The subsequent images are computed iteratively from the *seed* image using the semi-group property (1) to simulate the blurs following a geometric progression $\sigma_s = \sigma_{\min} 2^{s/n_{\text{spo}}}, s \geq 1$. The standard values proposed in [1] are $n_{\text{spo}} = 3$ and $\delta_{\min} = 1/2$.

The digital scale-space architecture is defined by four parameters: the number of octaves n_{oct}, the number of scales per octave n_{spo}, the initial oversampling factor $\delta_{\min}$, and the minimal blur level $\sigma_{\min}$ in the scale-space. Finally, the DoG scale-space is computed from the Gaussian scale-space as the difference between two successive images. The ratio between two successive blur levels is $\kappa = 2^{1/n_{\text{spo}}}$. Thus, by increasing n_{spo} the scale dimension can be sampled arbitrarily finely. In the same way by considering a small $\delta_{\min}$ value, the 2D position can be sampled finely.

3. THE THEORETICAL SCALE INVARIANCE

3.1. The camera model

In the present framework, the camera point spread function is modeled by a Gaussian kernel G_c and all digital images are frontal snapshots of an ideal planar object described by the infinite resolution image u_0. In the underlying SIFT invariance model, the camera is allowed to rotate around its optical axis, to take some distance, or to translate while keeping the same optical axis direction. All digital images can then be expressed as

$$\mathbf{u} =: \mathbf{S}_1 G_c H \mathcal{T} R u_0,$$

where $\mathbf{S}_1$ denotes the sampling operator, H a homothety, $\mathcal{T}$ a translation and R a rotation.

3.2. The SIFT method is invariant to zoom outs

It is not difficult to prove that SIFT is consistent with the camera model. Nevertheless, the proof in [6] is inexact, as pointed out in [36]. Let $\mathbf{u}_\lambda$ and $\mathbf{u}_\mu$ denote two digital snapshots of the scene u_0. More precisely,

$$\mathbf{u}_\lambda = \mathbf{S}_1 G_c H_\lambda u_0 \quad \text{and} \quad \mathbf{u}_\mu = \mathbf{S}_1 G_c H_\mu u_0.$$

Assuming that the images are well sampled and taking advantage of the semi-group property (1), the respective scale-spaces are

$$\begin{aligned} v_\lambda(\sigma, \mathbf{x}) &= G_{\sqrt{\sigma^2-c^2}} \mathbf{I}_1 \mathbf{S}_1 G_c H_\lambda u_0(\mathbf{x}) = G_\sigma H_\lambda u_0(\mathbf{x}), \\ v_\mu(\sigma, \mathbf{x}) &= G_\sigma H_\mu u_0(\mathbf{x}), \end{aligned}$$

where $\mathbf{I}_1$ denotes the interpolation operator. In fact, both scale-spaces only differ by a reparameterization. Indeed, if v_0 denotes the Gaussian scale-space of the infinite resolution image u_0 (i.e., $v_0(\sigma, \mathbf{x}) = G_\sigma u_0(\sigma, \mathbf{x})$) we have

$$\begin{aligned} v_\lambda(\sigma, \mathbf{x}) &= H_\lambda(G_{\lambda\sigma} u_0(\mathbf{x})) = v_0(\lambda\sigma, \lambda\mathbf{x}), \\ v_\mu(\sigma, \mathbf{x}) &= v_0(\mu\sigma, \mu\mathbf{x}), \end{aligned}$$

thanks to a commutation relation between homothety and convolution. With a similar argument, the two respective DoG functions are related to the DoG function w_0 derived from u_0. For a ratio $\kappa > 1$ we have

$$\begin{aligned} w_\lambda(\sigma, \mathbf{x}) &= v_\lambda(\kappa\sigma, \mathbf{x}) - v_\lambda(\sigma, \mathbf{x}) \\ &= v_0(\kappa\lambda\sigma, \lambda\mathbf{x}) - v_0(\lambda\sigma, \lambda\mathbf{x}) = w_0(\lambda\sigma, \lambda\mathbf{x}) \end{aligned}$$

and $w_\mu(\sigma, \mathbf{x}) = w_0(\mu\sigma, \mu\mathbf{x})$. Consider an extremum point $(\sigma_0, \mathbf{x}_0)$ of the DoG scale-space w_0. Then if $\sigma_0 \geq \max(\lambda c, \mu c)$, this extremum corresponds to extrema $(\sigma_1, \mathbf{x}_1)$ and $(\sigma_2, \mathbf{x}_2)$ of w_λ and w_μ respectively, satisfying $\sigma_0 = \lambda\sigma_1 = \mu\sigma_2$. This equivalence of extrema between the two scale-space guaranties that the SIFT descriptors are identical.

4. THE NUMERICAL SCALE INVARIANCE

To show how the scale invariance is affected by the scale-space sampling and the blur in the input image, we shall measure the invariance level by accurately simulating image pairs related through a scale change, a translation or a blur. We define the non repeatability ratio (NRR) as the number of keypoints detected in one image but not detected in the expected position of the other divided by the total number of detected keypoints. To define if a keypoint was correctly located, we used a more conservative tolerance than the classical one adopted by [11, 37]. We took an absolute tolerance of $\Delta x = \Delta y = 0.5$ px for the spatial position, and a relative tolerance of $\Delta s = 2^{1/4} s$ for the scale.

4.1. Simulating the digital camera

In our experiments, images were simulated to be accurately consistent with the SIFT camera model. Specifically, digital images were simulated from a large reference real digital image u_{ref} through Gaussian convolution and subsampling. To simulate a Gaussian camera blur c, a Gaussian convolution of standard deviation cS, with $S > 10$ was first applied. The convolved image was then subsampled by a factor S. Assuming that the reference image has an intrinsic Gaussian blur of $c_{\text{ref}} \ll cS$, the resulting Gaussian blur was $\sqrt{c^2 + (c_{\text{ref}}/S)^2} \approx c$. The blur level in natural images was

Fig. 1. Synthesized images `deer` and `pool` consistent with the image model. The respective blur levels are $c = 0.5$ and $c = 1.0$.

estimated from the point spread function of a consumer digital reflex camera following [38]. The obtained Gaussian blur levels varied from $c = 0.35$–0.95, depending on the aperture of the lens (blur level increases with the aperture size). Different zoomed-out and translated versions were simulated similarly by adjusting the scale parameter S and by translating the sampling grid.

Thanks to the large subsampling factor, the generated images could be considered to be noiseless. In addition, the images were stored with 32 bit precision to mitigate quantization effects. Figure 1 shows two of the simulated images used in the experiments.

It might be objected that our simulations are highly unrealistic as the images to be compared by SIFT in practice are not perfectly sampled or noiseless. Nevertheless, with an ever growing image resolution, more and more images will be compared after a large subsampling, so that these properties can become exact in practice. Furthermore, even if applying SIFT to the originals and regardless of initial noise and blur, the images at large scales also become anyway perfect so that the accuracy and repeatability issues under such favorable conditions are relevant.

4.2. The influence of scale-space sampling

We examined the detection stability when varying the number of scales per octave n_{spo} and the distance δ_{min}. Figure 3 shows the number of detected 3D extrema extracted from image `deer` using $n_{\text{spo}} = 2$–35 and $\delta_{\text{min}} = 1, 1/2, 1/4, 1/8, 1/16$. For a given spatial sampling rate, the number of detected extrema increases with n_{spo} and stabilizes for $n_{\text{spo}} > 15$. Setting $n_{\text{spo}} = 10$ and $\delta_{\text{min}} = 1/4$ gives a good trade-off between detection number and computational cost. Increasing the oversampling factor leads to a decrease of the number of detections which stabilizes for $\delta_{\text{min}} \leq 1/8$. The stabilization of the detection number seems to indicate that, once a sufficiently dense sampling is achieved, keypoint detection is stable.

By choosing a reference fine discretization, we were in a position to compare different configurations to check the stability of the detected keypoints. As a reference, we chose the keypoints detected with $n_{\text{spo}} = 24$, $\delta_{\text{min}} = 1/16$. We compare its detections to the ones obtained from coarser discretizations. Figure 2 shows that with coarse discretizations, SIFT fails to robustly detect the 3D extrema.

To examine the detection stability for different sampling parameters to image transformations, we considered a sub-pixel translation and a zoom-out. Figures 3 and 4 show the NRR and the number of detections for the translation and zoom-out respectively. The denser the sampling, the lower the NRR value, indicating that the extracted keypoints are more invariant to the transformations when the scale-space sampling is fine. In addition, the results show that it does not make sense to combine a high scale sampling rate with a low space sampling rate (or vice versa) as it leads to fewer invariant keypoints.

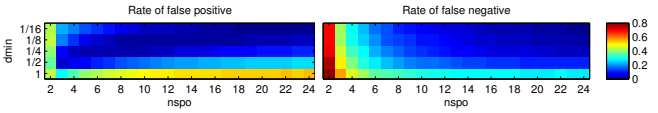

Fig. 2. Stability to different scale-space discretizations $(n_{\text{spo}}, \delta_{\text{min}})$. We considered as reference the keypoints detected from the finest discretization ($n_{\text{spo}} = 24$, $\delta_{\text{min}} = 1/16$). The left plot shows, as a function of the sampling parameters, the percentage of keypoints in coarser scale-spaces that are not invariant 3D extrema (i.e., not detected in the reference). The right plot shows the percentage of 3D extrema that are not detected in the coarser discretizations. All this indicates that SIFT fails to detect all 3D extrema unless the scale-space is significantly oversampled.

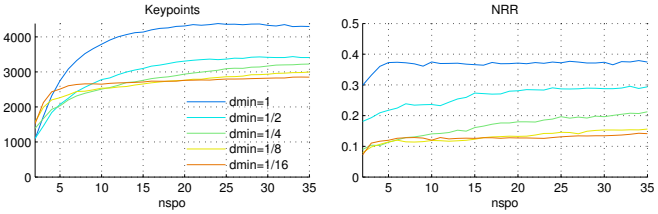

Fig. 3. The influence of scale-space discretization for a pair of translated images (`deer`, $c = 0.5$, translation of 0.25 px). On the left, the number of keypoints plotted as a function of the number of scales per octave n_{spo} for different spatial sampling rates δ_{min}. For a given δ_{min}, the number of detections increases with n_{spo} and stabilizes for $n_{\text{spo}} \geq 15$. For a given n_{spo}, the number of detections stabilizes for large oversampling factors ($\delta_{\text{min}} \leq 1/8$). The NRRs shown on the right plot indicates that keypoints detected with significantly oversampled scale-spaces are more stable to translation.

In conclusion, the standard setting of $n_{\text{spo}} = 3$, $\delta_{\text{min}} = 1/2$ is insufficient to robustly extract the 3D scale-space extrema. While the Gaussian scale-space may be well sampled according to the Nyquist rule, the rudimentary scanning for 3D extrema used in SIFT requires significant scale-space oversampling, e.g., $n_{\text{spo}} = 20$ and $\delta_{\text{min}} = 1/16$, to reliably detect all 3D extrema.

4.3. The influence of image blur

We also varied the input image blur c and examined how SIFT invariance is affected. Figures 5 and 6 show the number of detections and the NRR as a function of the minimal blur σ_{min} for the cases of a sub-pixel translation and a zoom-out respectively. The number of detections is the same regardless of the image blur. However, the NRR increases for lower values of c (see caption for details). The reason is that small c values produce undersampled sharp images that present aliasing artifacts generating non-invariant detections. The impact decreases for large σ_{min} but nevertheless stays noticeable in all octaves. On the other hand, for $c = 0.70$–1.10, the effect of image blur is not significant. Indeed, it is inexistent for $\sigma_{\text{min}} > 1.4$, which corresponds to structures larger than 3–4 pixels. As could be expected, SIFT performs better with smoothed-out well sampled images than with sharp aliased ones.

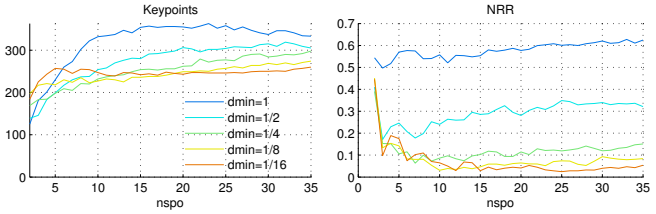

Fig. 4. The influence of scale-space discretization for a pair of images with different simulated zoom factor (`deer`, $c = 0.5$, relative zoom factor of 2.15). On the left, the number of keypoints in the zoomed-out image plotted as a function of the number of scales per octave n_{spo} for different spatial sampling rates δ_{min}. Oversampled scale-spaces lead to lower NRR values (shown on the right) which is an evidence of stability.

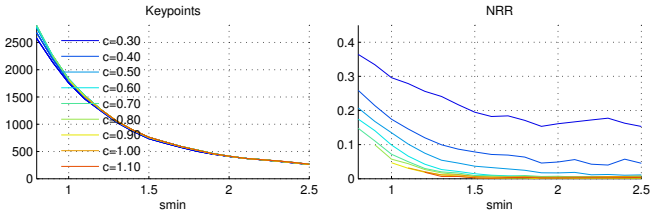

Fig. 5. The influence of image blur for a pair of translated images (`pool`, translation of 0.25 px). On the left, the number of keypoints plotted as a function of the minimal detection scale σ_{min} for different input image blur levels. Apart from the fact that no detection can be made below image blur ($\sigma_{\text{min}} \geq c$), the number of detections is the same regardless of the image blur. On the right, the NRR values plotted as a function of σ_{min} indicate that if the input image is undersampled ($c < 0.80$), aliasing will create non-invariant (spurious) detections. For $c = 0.30$–0.60, the impact of image blur decreases with σ_{min} but nevertheless stays noticeable in all octaves. While for $c = 0.70$–1.10, the impact of aliasing due to image blur is not significant, especially for $\sigma_{\text{min}} > 1.4$.

4.4. The DoG threshold

SIFT discards not well contrasted detections by using a threshold on the keypoint DoG values. To evaluate its effect, we applied a varying DoG threshold and examined if the surviving detections were more stable when considering different scale-space samplings and input image blurs. Figure 7 shows for two blur levels and two scale-space discretizations, the number of surviving detections and the NRR as a function of the DoG threshold for a subpixellically shifted image pair. This experiment proves that the elimination of keypoints resulting from the DoG threshold fails to improve the overall stability (see caption for details). We conclude that the unstable detections due to aliasing in the input image are well contrasted and cannot be discarded efficiently with the SIFT threshold.

5. CONCLUDING REMARKS

The above study demonstrates that the original parameter choice in SIFT is not sufficient to ensure a theoretical and practical scale invariance, which is the main claim of the SIFT method. The experiments also revealed that sharp images may deteriorate SIFT performance due to aliasing artifacts.

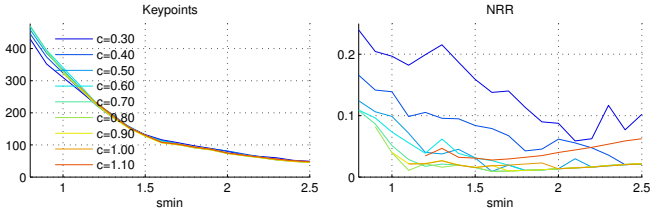

Fig. 6. The influence of image blur for a pair of images with different simulated zoom factors (`pool`, $c = 0.5$, relative zoom factor of 2.15). On the left, the number of keypoints in the zoomed-out image plotted as a function of σ_{min} for different input image blur levels. The NRR values as a function of σ_{min} are plotted on the right. High NRR values for low blur levels are explainable by unstable keypoints detected on aliased structures. Less sharp images lead to lower NRR values. The impact of image blur decreases with σ_{min}.

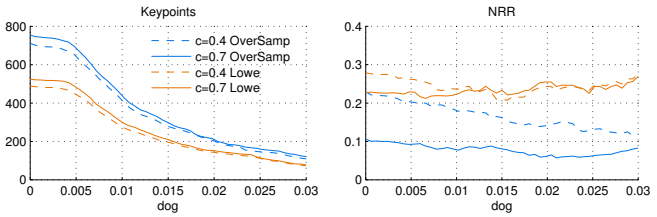

Fig. 7. Effect of the DoG threshold. We simulated a pair of translated images (`pool`, translation 0.25 px) with two image blurs $c = 0.4, 0.7$. SIFT was applied with two scale-space discretizations: the reference ($n_{\text{spo}} = 3, \delta_{\text{min}} = 1/2$) denoted `Lowe` and an oversampled scale-space ($n_{\text{spo}} = 30, \delta_{\text{min}} = 1/16$) denoted `OverSamp`. On the left, the number of surviving detections as a function of the DoG threshold. On the right, the NRRs as a function of the DoG threshold. The DoG threshold fails to significantly improve the overall stability of keypoints.

Our scope was not to propose a new or optimized SIFT. Nevertheless, some practical conclusions can be drawn from our observations. The repeatability curves for an oversampled SIFT show that a $4\times$ space oversampling (instead of 2) and a $10\times$ scale oversampling (instead of 3) ensure a twice lower non-repeatability and twice more keypoints. There is no question that this detection/repeatability improvement is desirable. The main objection is its computational cost, which is multiplied by 7 per detected keypoint. Yet, this increased computational expense affects only the detection phase. The found descriptors are more repeatable and therefore better. It follows that the overall efficiency of the method is increased at fixed cost per image. Thus when matching an image to a large descriptor database, this oversampling is preferable, as the main computational cost is for descriptor comparison. Furthermore, the complexity objection does not apply to the keypoint comparison after the third octave, when JPEG, aliasing and noise artifacts are minimal and therefore the subsampled images are almost perfect.

In short, a significantly more invariant SIFT can be made by simply oversampling in scale and space after the third octave for normal images, and by oversampling from the first scale for good quality uncompressed images.

The DoG was originally conceived as an approximation of the the Laplacian of Gaussian. However, this is not necessarily true and will be the object of future research. Finally, the present analysis did not tackle image noise and an uncertainty in the input image blur. These are left as future work.

6. REFERENCES

[1] D. Lowe, "Object recognition from local scale-invariant features," in *ICCV*, 1999.

[2] D. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, pp. 91–110, 2004.

[3] M. Brown and D. Lowe, "Automatic panoramic image stitching using invariant features," *IJCV*, vol. 74, no. 1, pp. 59–73, 2007.

[4] F. Riggi, M. Toews, and T. Arbel, "Fundamental matrix estimation via TIP-transfer of invariant parameters," in *ICPR*, 2006.

[5] C. Strecha, W. von Hansen, L. Van Gool, P. Fua, and U. Thoennessen, "On benchmarking camera calibration and multi-view stereo for high resolution imagery," in *CVPR*, 2008.

[6] J.-M. Morel and G. Yu, "Is SIFT scale invariant?," *Inverse Problems and Imaging*, vol. 5, no. 1, pp. 115–136, 2011.

[7] J. Weickert, S. Ishikawa, and A. Imiya, "Linear scale-space has first been proposed in Japan," *J. Math. Imaging Vision*, vol. 10, no. 3, pp. 237–252, 1999.

[8] T. Lindeberg, *Scale-space theory in computer vision*, Springer, 1993.

[9] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Found. Trends in Comp. Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008.

[10] H. Bay, T. Tuytelaars, and L. van Gool, "SURF: Speeded Up Robust Features," in *ECCV*, 2006.

[11] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, "A comparison of affine region detectors," *IJCV*, vol. 65, no. 1-2, pp. 43–72, 2005.

[12] W. Förstner, T. Dickscheid, and F. Schindler, "Detecting interpretable and accurate scale-invariant keypoints," in *ICCV*, 2009.

[13] P. Mainali, G. Lafruit, Q. Yang, B. Geelen, L. Van Gool, and R. Lauwereins, "SIFER: Scale-Invariant Feature Detector with Error Resilience," *IJCV*, vol. 104, no. 2, pp. 172–197, 2013.

[14] C. Ancuti and P. Bekaert, "SIFT-CCH: Increasing the SIFT distinctness by color co-occurrence histograms," in *ISPA*, 2007.

[15] O. Pele and M. Werman, "A linear time histogram metric for improved sift matching," in *ECCV*. 2008.

[16] J. Rabin, J. Delon, and Y. Gousseau, "A statistical approach to the matching of local features," *SIAM J. Imaging Sci.*, vol. 2, no. 3, pp. 931–958, 2009.

[17] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *CVPR*, 2004.

[18] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary Robust Independent Elementary Features," in *ECCV*. 2010.

[19] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *ICCV*, 2011.

[20] E. Tola, V. Lepetit, and P. Fua, "A fast local descriptor for dense matching," in *CVPR*, 2008.

[21] E. Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *PAMI*, vol. 32, no. 5, pp. 815–830, 2010.

[22] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," in *Proc. ACM Int. Conf. Multimed.*, 2010.

[23] S. Leutenegger, M. Chli, and R.Y. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints," in *ICCV*, 2011.

[24] M. Agrawal, K. Konolige, and M.R. Blas, "CenSurE: Center Surround Extremas for Realtime Feature Detection and Matching," in *ECCV*. 2008.

[25] S. Winder and M. Brown, "Learning local image descriptors," in *CVPR*, 2007.

[26] S. Winder, G. Hua, and M. Brown, "Picking the best DAISY," in *CVPR*, 2009.

[27] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao, "WLD: A robust local image descriptor," *PAMI*, vol. 32, no. 9, pp. 1705–1720, 2010.

[28] M. Grabner, H. Grabner, and H. Bischof, "Fast approximated SIFT," in *ACCV*. 2006.

[29] C. Liu, J. Yuen, A. Torralba, J. Sivic, and W.T. Freeman, "SIFT Flow: Dense correspondence across different scenes," in *ECCV*. 2008.

[30] P. Moreno, A. Bernardino, and J. Santos-Victor, "Improving the SIFT descriptor with smooth derivative filters," *Pattern Recognition Lett.*, vol. 30, no. 1, pp. 18–26, 2009.

[31] M. Brown, R. Szeliski, and S. Winder, "Multi-image matching using multi-scale oriented patches," in *CVPR*, 2005.

[32] T. Dickscheid, F. Schindler, and W. Förstner, "Coding images with local features," *IJCV*, vol. 94, no. 2, pp. 154–174, 2011.

[33] R. Sadek, C. Constantinopoulos, E. Meinhardt, C. Ballester, and V. Caselles, "On affine invariant descriptors related to SIFT," *SIAM*, vol. 5, no. 2, pp. 652–687, 2012.

[34] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *PAMI*, vol. 27, no. 10, pp. 1615–1630, 2005.

[35] K.E.A. Van De Sande, T. Gevers, and C.G.M. Snoek, "Evaluating color descriptors for object and scene recognition," *PAMI*, vol. 32, no. 9, pp. 1582–1596, 2010.

[36] R. Sadek, *Some problems on temporally consistent video editing and object recognition*, Ph.D. thesis, Universitat Pompeu Fabra, 2012.

[37] K. Mikolajczyk and C. Schmid, "Scale & affine invariant interest point detectors," *IJCV*, vol. 60, no. 1, pp. 63–86, 2004.

[38] M. Delbracio, P. Musé, and A. Almansa, "Non-parametric subpixel local point spread function estimation," *IPOL*, 2012.